Armadillo-repeat protein functions: questions for little creatures

Tewari, R; Bailes, E; Bunting, KA; Coates, Juliet

DOI:
10.1016/j.tcb.2010.05.003

Document Version
Early version, also known as pre-print

Citation for published version (Harvard):
Tewari, R, Bailes, E, Bunting, KA & Coates, J 2010, 'Armadillo-repeat protein functions: questions for little creatures', *Trends in Cell Biology*, vol. 20, no. 8, pp. 470-481. https://doi.org/10.1016/j.tcb.2010.05.003

Review

Cell PRESS

Armadillo-repeat protein functions: questions for little creatures

Rita Tewari[1], Elizabeth Bailes[1], Karen A. Bunting[1] and Juliet C. Coates[2]

[1] Institute of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK
[2] School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Armadillo (ARM)-repeat proteins form a large family with diverse and fundamental functions in many eukaryotes. ARM-repeat proteins have largely been characterised in multicellular organisms and much is known about how a subset of these proteins function. The structure of ARM-repeats allows proteins containing them to be functionally very versatile. Are the ARM-repeat proteins in 'little creatures' as multifunctional as their better-studied relatives? The time is now right to start analysing ARM-repeat proteins in these new systems to better understand their cell biology. Here, we review recent advances in understanding the many cellular roles of both well-known and novel ARM-repeat proteins.

Armadillo-repeat containing proteins

Armadillo (ARM)-repeat proteins are characterised by containing a repeating ~42 amino acid motif composed of three α-helices, which was first characterized in the *Drosophila* segment polarity protein Armadillo [1]. Several ARM-repeat protein crystal structures have been solved [2–11], demonstrating that although ARM-repeat proteins do not necessarily share a great deal of sequence identity (e.g. [9,12]), they share a related structure (Figure 1) and are evolutionarily ancient (Box 1). Tandem ARM-repeat units fold together as a superhelix, forming a versatile platform for interactions with many protein partners. For this reason, many ARM-repeat proteins have more than one independent cellular role, and eukaryotic ARM-repeat proteins as a whole have diverse and important functions [13]. The lack of shared sequence identity and the degenerate nature of the repeat sequence (Figure 1) makes defining cross-species homologues and orthologues of ARM-repeat proteins problematic, although ARM-repeat detection methods are improving [14]. Several ARM-repeat proteins have been identified solely based on solution of their crystal structure (e.g. [6,7,9]), whereas whole-genome sequencing projects have enabled the annotation of many putative ARM-repeat protein homologues ripe for study, throughout the tree of life (Figures 2 and 3; Boxes 2 and 3), particularly in important unicellular eukaryotes.

The era of large-scale genome sequencing has enabled identification of many putative ARM-repeat proteins throughout the tree of life, particularly in unicellular organisms such as disease-causing protists and algae. This has provided a goldmine of new data for eukaryotic cell biologists, which challenges some old assumptions and highlights novel systems for future research.

β-catenin/Armadillo: the prototypical ARM-repeat protein

β-catenin (Armadillo in *Drosophila*) is a fascinating protein with many important cellular and developmental functions. The roles of β-catenin are 'classically' defined: (i) as an adhesion protein and (ii) as a signalling protein, transducing extracellular signals to the nucleus to modify gene expression. β-catenin has many binding partners that mediate a diverse set of cellular functions, and the protein probably acts as a 'hub' on which many cellular signalling networks impinge.

β-catenin is a key node in Wnt signalling throughout the animal kingdom [15,16]. Until cells receive a Wnt signal, β-catenin is maintained in an unstable state by the concerted action of several kinases and scaffold proteins (referred to as the cytosolic 'destruction complex' [15]). Stabilised β-catenin enters the nucleus where it binds to transcription factors, including those of the Lymphoid Enhancer Factor/T-Cell Factor (LEF/TCF) family, and hence turns key developmental- and cell proliferation genes on or off [15,16]. The structure of β-catenin is key to its regulation during Wnt signalling: many β-catenin interaction partners bind to a positively-charged groove in the ARM-repeat region ([17]; Figure 1).

Mammalian β-catenin was originally discovered as a component of actin-containing junctions that link cells together via cadherin proteins. It is now thought that cell junctions are dynamic structures that regulate actin dynamics locally to the junction [18]. The cytoplasmic tail of cadherin molecules binds to the same area of the β-catenin ARM-repeats as various Wnt signalling components, thus providing a means for crosstalk between cell junctions and cell signalling pathways, since binding of cadherin and signalling components is mutually exclusive [17].

β-catenin also interacts with microtubules, including by localising to centrosomes and regulating their regrowth, cohesion and separation during mitosis [19–21]. Centrosome splitting is promoted by Wnt signals [20], thus demonstrating both a transcriptional and structural role for Wnt/β-catenin signalling during cell proliferation. β-catenin also contacts the cytoskeleton indirectly via its interaction with Adenomatous Polyposis Coli (APC), a large multi-domain protein that is part of the β-catenin destruction complex described above [22]. Interestingly,

Corresponding author: Coates, J.C. (j.c.coates@bham.ac.uk).

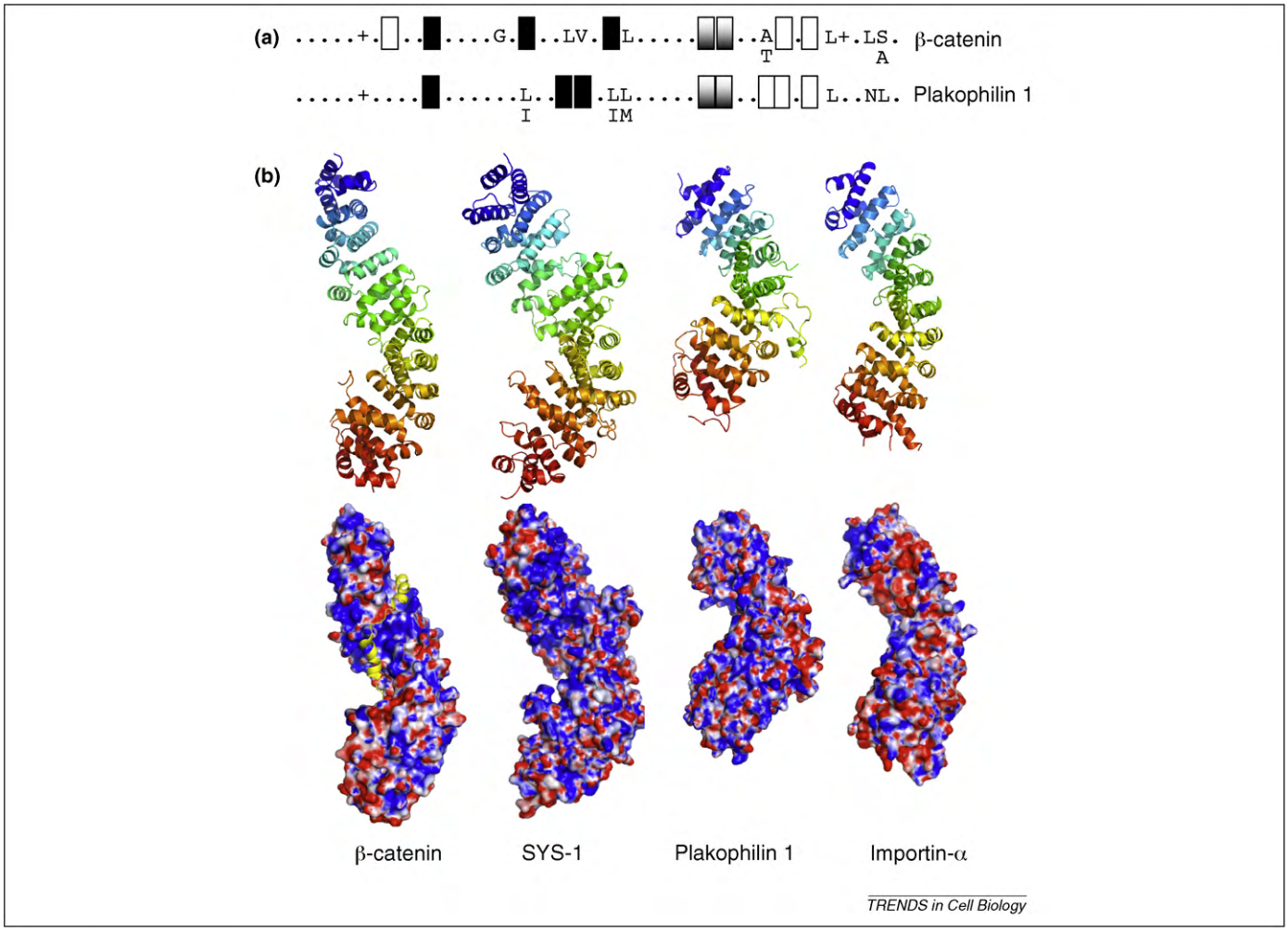


Figure 1. Sequences and structures of ARM-repeat proteins. **(a)** Consensus single ARM-repeat amino-acid sequences for β-catenin [4] and Plakophilin [2]. Note the lack of conservation at many positions in the repeat, leading to potentially low sequence similarities in different ARM-repeats. +, basic (H,K,R); white square, small hydrophobic (ACPVT); black square, large hydrophobic (FILMW); two-tone square, any hydrophobic. **(b)** Conserved structures of the ARM-repeat regions of different ARM-repeat proteins. Top, ribbon diagrams and bottom, electrostatic surfaces for (from left to right) Human β-catenin (12 ARM-repeats: 1JDH in Protein Data Bank), *C. elegans* SYS-1 (12 ARM-repeats: 3C2H), Plakophilin 1 (10 ARM-repeats; 1XM9) and importin-α (10 ARM-repeats; 1BK5). The N-terminal end of the protein is at the top in each representation. In the lower diagram, the accessible surface area is coloured according to electrostatic potential from −10 k_BT/e (red) to + 10 k_BT/e (blue). The electrostatic potential was calculated using APBS [139]. The TCF4 peptide (yellow) complexed with human β-catenin is shown in cartoon representation, binding in the positively charged groove in the ARM-repeats. Note importin-α has very different overall charge distribution but similar structure to the other proteins. Figures were produced using PyMol [140].

APC itself also contains ARM-repeats, which themselves interact with various regulators and effectors of small GTPases that control cell migration via the actin cytoskeleton [22]. APC also interacts with microtubules via several different domains ([22]; see below). This enables control of cell migration (with additional crosstalk to the actin cytoskeleton), as well as chromosome separation during cell division [22].

The identified functional repertoire of β-catenin is ever-growing. It has diverse interaction partners in several subcellular compartments that bind to its ARM-repeats (e.g. http://www.stanford.edu/~rnusse/pathways/binding.html). Various tissue-restricted proteins bind to the ARM-repeat region of β-catenin to regulate its stability, including the Wilms' Tumour suppressor protein WTX [23], the renal tumour suppressor JADE-1 [24] and the DNA damage-induced SIAH-1 [25].

β-catenins are thought to function similarly throughout the animal kingdom. β-catenin signalling and junctional localisation are present in cnidarians, the earliest-evolving radially symmetrical animals [26], and β-catenins are present in sponges, the phylogenetically most basally branching animals [27,28]. The dual adhesion/signalling functions are proposed to have been present ~700 million years ago, in the last common ancestor of animal β-catenin [29]. In zebrafish, genome duplication has resulted in two β-catenins with overlapping but distinct functions [30]. In *Caenorhabditis elegans*, the situation is more extreme, with four highly divergent β-catenins having evolved distinct roles in adhesion (HMP-2) or signalling (BAR-1, WRM-1, SYS-1) [31].

Proteins related to β-catenin

Animals possess several proteins related in function to β-catenin. Vertebrate plakoglobin arose from a β-catenin gene duplication in the chordate lineage [29]; human β-catenin and plakoglobin are 69% identical. Plakoglobin substitutes for β-catenin function in some instances, but

Box 1. Importin-α: fundamental to eukaryotic cells

Importin-α transports cargo proteins into the nucleus and consists of 10 ARM-repeats downstream of a structurally related importin-β-binding (IBB) domain ([13]; Figure 3). There are three subclasses of animal importin-α. Ancestral importin-α genes gave rise to the α1 subclass, which is also present in plants, fungi, amoebae and choanoflagellates [117]. Animal importin-α2 and -α3 proteins have acquired unique functions during development and differentiation processes, particularly during gametogenesis [117–120]. In *Drosophila*, mouse and *C. elegans* various importin-α genes are specifically expressed during spermatogenesis or gametogenesis and the absence of these genes leads to male and female sterility in *Drosophila* [119,121]. Similarly in plants, the ancestral importin-α1-like gene has diversified to give plant-specific importin-α proteins, some of which have specific cellular roles [122,123].

Is importin-α the oldest and perhaps ancestral ARM-repeat protein? Importin-α is conserved across all eukaryotes, highlighting its ancient nature [117,124]. The nucleus defines a eukaryotic cell and is clearly an ancient organelle, which probably arose from an endosymbiotic event. Members of the nuclear import and export machinery are all structurally related to one another and probably shared a common prokaryotic ancestor, raising interesting questions about the sequential evolution of the nuclear transport machinery [124]. It is thought that ARM-repeat proteins evolved from duplication of a single ARM-repeat structure within an ancestral protein, and that tandem ARM-repeats were present in the last eukaryotic common ancestor [125].

It seems possible that other ARM-repeat proteins evolved from importins or an importin-like ancestor. Like importin-α, many ARM-repeat proteins are able to enter the nucleus independently of a nuclear localisation signal [37,53,126]. Additionally, importin-α, like β-catenin and related proteins, interacts with and possibly regulates both actin [13] and microtubules [127]. Our searches have uncovered many unknown proteins simply annotated as 'importin-α-like' in diverse unicellular eukaryotes and ascertaining the functions of these proteins might shed light onto the evolution of ARM-repeat proteins and their function.

has also evolved unique functions: it is a component of actin-containing cell–cell junctions, but unlike β-catenin, it is also found in desmosomes, specialised intermediate-filament-containing junctions that protect tissues from mechanical stress. Plakoglobin mutations are embryolethal and particularly affect the heart, which is under particular stress, thus rich in desmosomes. Plakoglobin also binds to LEF/TCF proteins in the nucleus: whether plakoglobin acts redundantly with or antagonistically to β-catenin might depend on the cell type being analysed [32–34].

The p120 family of proteins [p120 catenin, δ-catenin, p0071 and Armadillo-Repeat gene deleted in Velo-Cardio-Facial syndrome (ARVCF)] is found in vertebrates, *Drosophila* and *C. elegans* [35]. p120 proteins, like β-catenin, bind to classical cadherins using their ARM-repeats. p120s facilitate microtubule-based trafficking of cadherins, stabilise cadherin clusters at the cell surface and also block cadherin endocytosis and degradation, thus controlling the strength of cell–cell adhesion by regulating the number of cadherin molecules at the plasma membrane [35].

p120 catenin (referred to here simply as p120) regulates the actin cytoskeleton via small GTPases, inhibiting RhoA activity and activating Rac and Cdc42. p120 binding to cadherins or RhoA is mutually exclusive as both interactions involve the ARM-repeats. Activation of Rac by p120 antagonises the association of p120 (indirect, involving the ARM-repeats) with microtubules [36]. Thus p120 could mediate a balance between motility and adhesion by several mechanisms [35].

Like β-catenin and plakoglobin, p120 is present in the nucleus, where its ARM-repeats interact with the transcription factor Kaiso [37]. This interaction is required for *Xenopus* gastrulation, in particular convergent-extension cell movements mediated by 'non-canonical' Wnt signalling (i.e. Wnt signalling that does not involve β-catenin; [38,39]). δ-catenin, like β-catenin, is subject to phosphorylation and proteasomal degradation [40].

Importantly, Kaiso also binds to LEF/TCF family transcription factors and inhibits their activity; this inhibition is relieved by p120. Thus, β-catenin and p120 together activate at least some 'canonical' Wnt-induced gene expression [39,41,42]. Interestingly, some p120/Kaiso-mediated transcriptional regulation could also involve interaction between p120 and a nuclear-localised fragment of the E-cadherin C-terminus [43].

The vertebrate-specific plakophilin family is most closely related to p120s: human plakophilins are between 25% and 47% identical to human p120 proteins. Plakophilins regulate junctional assembly, strength and crosstalk; like other ARM-repeat proteins they have multiple interaction partners [44,45].

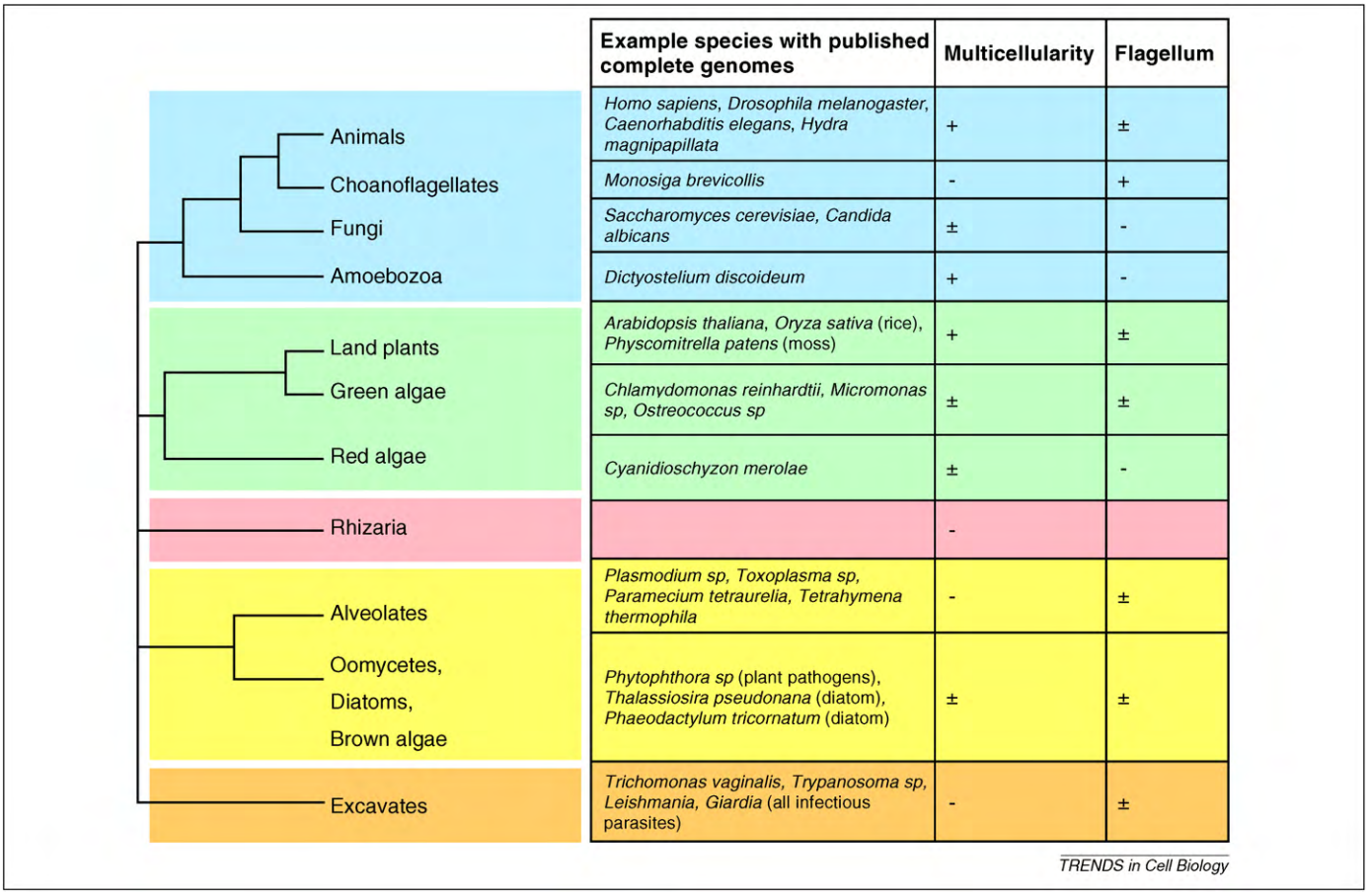
β-catenin-like proteins outside the animal kingdom

Do proteins with the same functions as β-catenin exist outside multicellular animals? Many proteins and protein domains associated with animal cell–cell adhesion are present widely throughout eukaryotes, including in unicells [46], so this seems a reasonable question to pose. Indeed, there are some candidate proteins that have some level of shared sequence identity with β-catenin.

The Vac8p protein [47] in the unicellular yeast *Saccharomyces cerevisiae* (Figure 2) is 22% identical to human β-catenin. Vac8p has several vacuolar functions, is a component of an actin-containing intra-organellar junction between the vacuole and the nucleus, and has a novel role in caffeine resistance [13,48]. Vac8p has unique binding partners for each of its functions, which target Vac8p to distinct sub-regions of the vacuolar membrane and bind in a mutually exclusive manner to the ARM-repeat domain [48]. The Vac8p homologue in *Candida albicans*, a fungus that exhibits multicellular (hyphal) growth under starvation conditions, is required for vacuolar inheritance and for regulation of hyphal branching frequency, co-ordinating vacuolar inheritance with cell size and the cell cycle [49,50].

Although Vac8p has a role in actin-containing intracellular junctions, it does not share all the functions of β-catenin [48]. Vac8p currently has no known cell-signalling role. However, as with β-catenin, the ARM-repeat region of Vac8p provides a platform for interaction with many protein partners, leading to the functional versatility of Vac8p.

If Vac8p and β-catenin proteins shared a common ancestor in the ancestor of animals and fungi (Figure 2) the proteins have clearly diverged in function to a great degree. Choanoflagellates are the unicellular organisms most closely related to animals [51]. No β-catenin has been found in the choanoflagellate *Monosiga* (Figure 2); the cadherins of



Example species with published complete genomes	Multicellularity	Flagellum
Homo sapiens, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Hydra magnipapillata*	+	±
Monosiga brevicollis	-	+
Saccharomyces cerevisiae, *Candida albicans*	±	-
Dictyostelium discoideum	+	-
Arabidopsis thaliana, *Oryza sativa* (rice), *Physcomitrella patens* (moss)	+	±
Chlamydomonas reinhardtii, *Micromonas sp*, *Ostreococcus sp*	±	±
Cyanidioschyzon merolae	±	-
	-	
Plasmodium sp, *Toxoplasma sp*, *Paramecium tetraurelia*, *Tetrahymena thermophila*	-	±
Phytophthora sp (plant pathogens), *Thalassiosira pseudonana* (diatom), *Phaeodactylum tricornatum* (diatom)	±	±
Trichomonas vaginalis, *Trypanosoma sp*, *Leishmania*, *Giardia* (all infectious parasites)	-	±

Figure 2. The eukaryotic tree of life. Simplified representation of the eukaryotic tree of life (adapted from [141]) showing the five major kingdoms. Blue: unikonts, green: green and red algae and plants, pink: rhizaria (no genome sequences available); yellow, chromoalveolates {consisting of two major lineages: **(i)** alveolates (including Apicomplexans – Box 2) and **(ii)** oomycetes, diatoms and brown algae [141]; orange: excavates. Examples of model organisms with sequenced genomes in each group are shown where applicable. Fundamental cell biological properties for each group (discussed in the main text, i.e. existence of multicellular forms, presence of a flagellum) are shown where known: +, present; −, absent; ±, present in some species only.

Monosiga lack a β-catenin binding domain [52]. We do not currently know whether β-catenin was lost from choanoflagellates, or if it was never present in the ancestor of choanoflagellates and multicellular animals. The evolution of β-catenin might have been a prerequisite for animal multicellularity, perhaps by providing the necessary 'missing link' between cadherins and the actin cytoskeleton that enabled stable cell–cell contacts to form.

The social amoeba *Dictyostelium* (Figure 2) has a β-catenin-related protein, Aardvark, which localises to actin-containing cell-cell junctions and affects cell differentiation [13]. Aardvark is 18% identical to human β-catenin, a similar level of identity as between human β-catenin and the *C. elegans* β-catenin homologue WRM-1 [31]. Aardvark contains an F-box motif, as do the related Arabidillo proteins in the land plant *Arabidopsis*, which regulate development, in particular root branching ([13,53]; Figure 3). F-box proteins are components of E3 ubiquitin ligases [54], suggesting a possible protein degradation function for Aardvark and Arabidillos, different from β-catenin functions, although this has not yet been confirmed. The *Arabidopsis* ARM-Repeat Only protein ARO1 regulates cell growth and actin organisation [55]. ARO1 partially co-localises with F-actin, but is also seen in nuclei and at the plasma membrane, suggesting that, like β-catenin, it is a multifunctional protein [55].

Other ARM-repeat proteins and the cytoskeleton

The cytoskeleton is a fundamental component of all eukaryotic cells. As described above, animal β-catenin and p120 and yeast Vac8p all associate with actin, microtubules or both. In the next sections we focus on ARM-repeat proteins with putative homologues throughout the tree of life (Figures 2 and 3), which could have conserved cytoskeletal functions awaiting discovery in a variety of important unicellular eukaryotes.

Kinesin-associated protein 3 (KAP3)

Kinesin-Associated Protein 3 (KAP3) is a multifunctional ARM-repeat protein within a motor protein complex that moves cargo along microtubules, particularly in neuronal axons and flagella [56]. KAP3 participates in MAP-kinase signalling and chromosome segregation during mitosis [57] and interacts with other ARM-repeats, both in APC and in the nucleotide exchange factor SMG-GDS/VIMAR [13]. In mammals, KAP3 might control neurotransmitter release [58,59] and human homologues are implicated in degenerative motor neurone diseases [60].

Our similarity searches reveal proteins related to KAP3 encoded in the genomes of *Monosiga* [61], in unicellular green algae [62,63], in a diatom [64] and in various chromoalveolate and excavate parasites [65–69]. Protein sequence identity to human KAP3 (GenBank accession

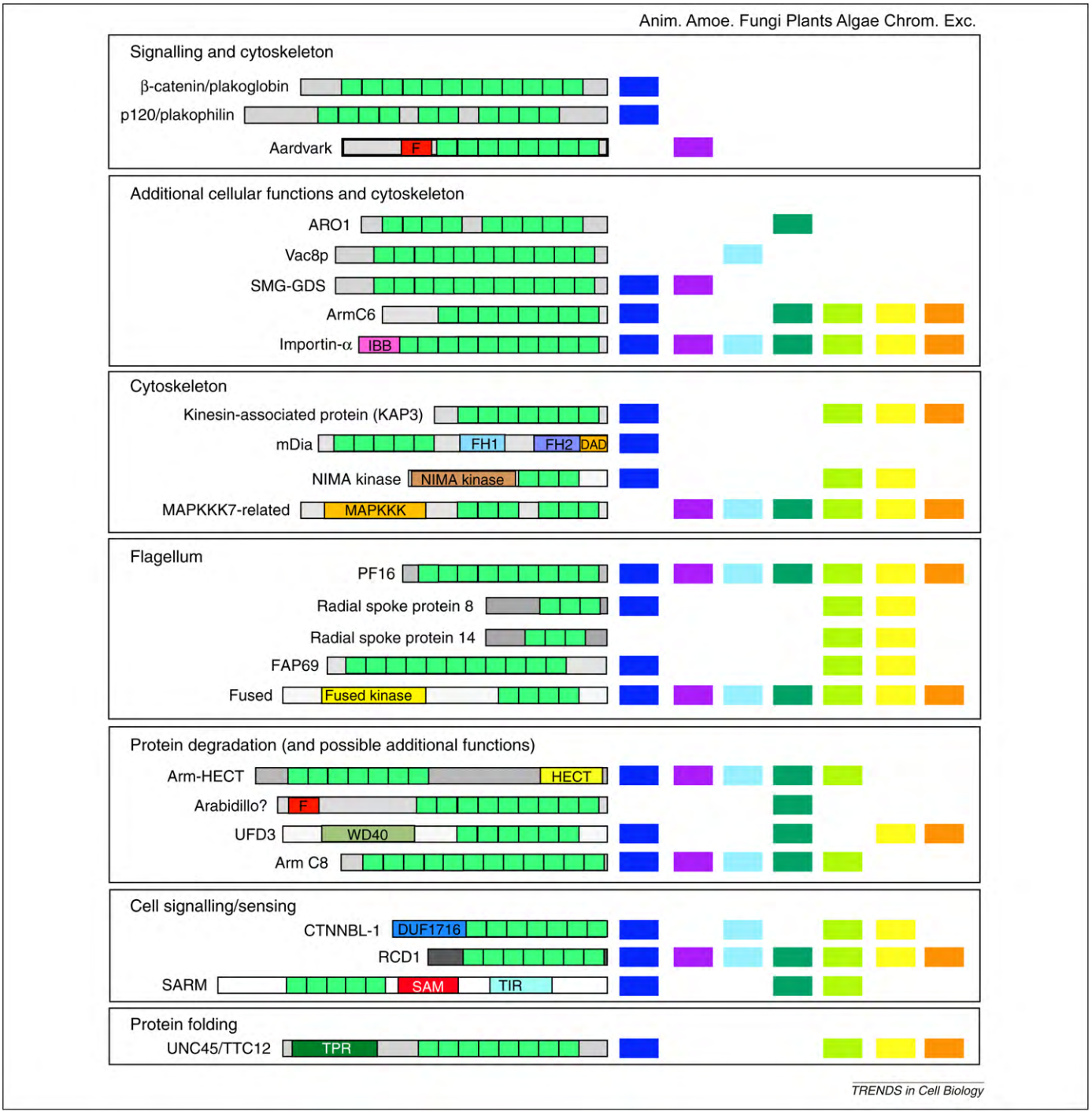


Figure 3. ARM-repeat proteins: functions and conservation. Representation of ARM-repeat proteins discussed in the text that have been identified in more than one species. The presence of each type of protein in various branches of the eukaryotic tree of life is indicated by coloured boxes: animals (Anim. – dark blue boxes), amoebae (Amoe. – purple boxes), fungi (light blue boxes), plants (dark green boxes), green/red algae (Algae – light green boxes), Chromoalveolates (Chrom. – yellow boxes,), Excavates (Exc. – orange boxes) are shown. Proteins are grouped according to function as in the main text. ARM-repeats are shown in mid-green. Protein domain abbreviations: F, F-box; IBB, importin beta binding; FH1, Formin Homology 1; FH2, Formin Homology 2; DAD, Diaphonous Autoregulatory Domain; NIMA; Never In Mitosis A; MAPKKK, Mitogen Activated Protein-kinase-kinase-kinase; HECT, Homology to E6 C-terminus; WD40, WD40 repeats; DUF1716, domain of unknown function 1716; SAM, sterile alpha motif; TIR, Toll-like interleukin receptor; TPR, tetratricopeptide repeats.

AL121714) ranges from 18% (hypothetical protein in *Trichomonas vaginalis*, [65]) to 45% (hypothetical protein in *Monosiga brevicollis*, [61]). Thus, the KAP3 ancestor could have been present in the ancestor of all eukaryotes. No putative KAP3 has been detected in the sequenced genomes of higher plants or *Dictyostelium*. However, SMG-GDS is found in *Dictyostelium*, where it is required for chemotaxis but not cell division [13]. *Dictyostelium* (a non-flagellated organism) might have lost KAP3 [70]. We can speculate that KAP3 acquired a novel role interacting with SMG-GDS specifically in the animal lineage. Given its roles in cell transport and cell division in multicellular organisms, it will be interesting to ascertain what key functions KAP3 has in unicells from diverse kingdoms.

Box 2. Apicomplexans: parasites with novel ARM-repeat proteins

The Apicomplexa is a phylum comprising over 5000 unicellular parasites that are important pathogens affecting humans and livestock [128]. Notable Apicomplexa include *Plasmodium*, the causative agent of malaria; *Toxoplasma*, causing toxoplasmosis; *Cryptosporidium*, a waterborne diarrhoea-causing pathogen; *Theileria*, causing East Coast fever; and *Babesia*, causing bovine tick fever. Apicomplexa contain an essential apicoplast, a relic non-photosynthetic plastid from engulfment and secondary endosymbiosis of a red alga. In addition all Apicomplexa possess an apical complex at the anterior end of the cell, which contains specialised vesicles that secrete enzymes required for host cell invasion.

The Apicomplexan life cycle is complex, and has diverged in different parasites. *Cryptosporidium* has the simplest life cycle, invading a single cell type (intestinal epithelium) in a single host (human). *Plasmodium*, *Theileria* and *Babesia* undergo multiplication and asexual reproduction in the red blood cells of their vertebrate host(s) but require an insect vector for their sexual reproduction and transmission. *Plasmodium* species affecting mammals are transmitted by the *Anopheles* mosquito, whereas *Plasmodium gallinaceum*, affecting birds, is transmitted by *Aedes*. *Theileria* and *Babesia* have a tick vector [129–131].

Because of their medical and veterinary importance the genomes of several Apicomplexa have recently been sequenced [132]. Genome sizes vary widely: *Plasmodium* species have 14 chromosomes and genomes of 23–30 Mb whereas *Theileria* and *Babesia* have four chromosomes totalling 8–10 Mb [129]. Importantly, some Apicomplexan species are genetically tractable (e.g. *Plasmodium* and *Toxoplasma*) and amenable to the study of gene function and cell biology [133,134].

By mining Apicomplexan genomes using ApiDB and PlasmoDB [132], we have identified ten putative Apicomplexan ARM-repeat proteins (Figure I). Half of these proteins are conserved throughout eukaryotes and their functions are described in the main text. Interestingly, PF16 appears to be absent from *Theileria*, *Cryptosporidium* and *Babesia* (although incomplete genome annotation is possible). This could be owing to differences in gamete biology: *Plasmodium* and *Toxoplasma* have flagellated male gametes, whereas *Theileria*, *Cryptosporidium* and *Babesia* lack any flagellated stage to their life cycle [135,136].

The remaining Apicomplexan ARM-repeat proteins we have found are novel. The absence of two of these (ACU12396, AAN37153) in *Cryptosporidium* could reflect its simpler life cycle and host range. We hypothesise that Apicomplexan-specific ARM-proteins are important for key aspects of the biology of these parasites. Thus, these proteins and their putative signalling pathways are candidates for novel therapeutic targets in the future.

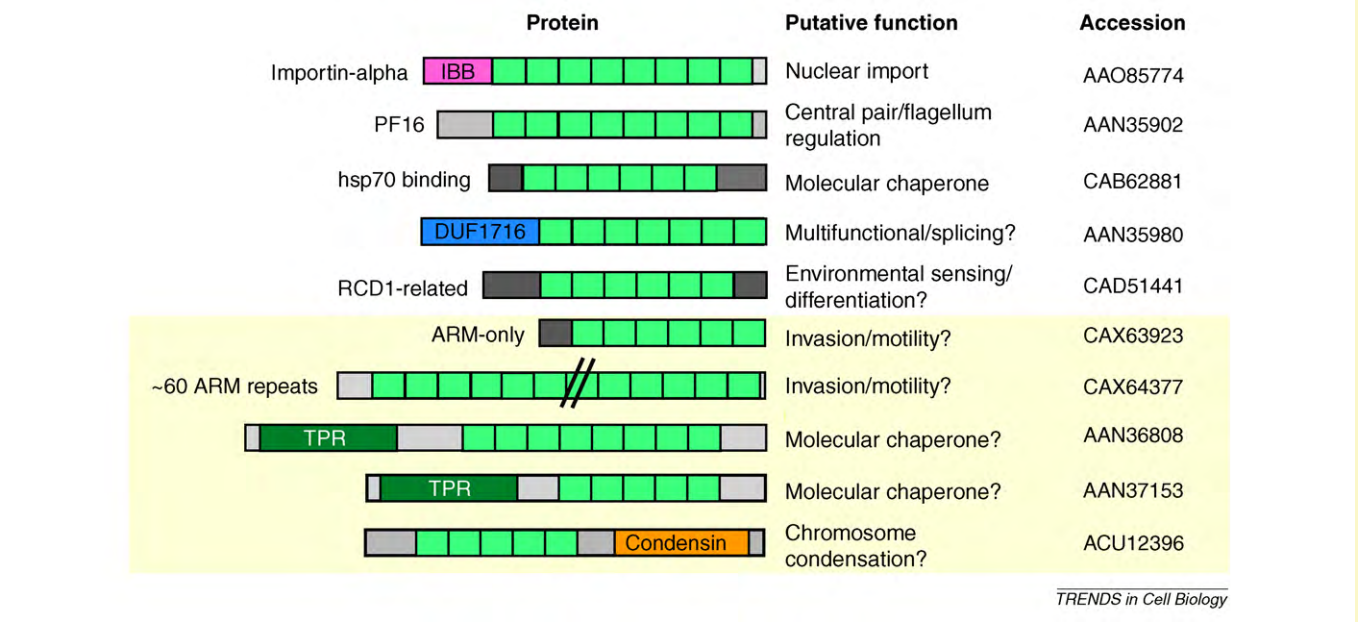


Figure I. ARM-repeat protein families in Apicomplexans. Domain structures and putative functions for each protein are given. ARM-repeats are shown in green. Protein abbreviations: IBB, importin-β binding; DUF1716, domain of unknown function 1716; TPR, tetratricopeptide repeats. The five Apicomplexan-specific ARM-repeat proteins are highlighted with yellow shading. Representative accession numbers are given for *Plasmodium falciparum*. Note that *Toxoplasma* might have more than one close homologue of some of these proteins.

Diaphonous-related formins: a novel ARM-repeat structure

Formins are found in animals, plants and fungi, where they nucleate and bundle actin filaments and stabilise microtubules during cell growth, adhesion, differentiation, cytoplasmic motility and cytokinesis [71,72]. Mammalian Diaphonous-related formins (DRFs) are unique formins containing ARM-repeats [6,7]. DRFs are activated by Rho-family GTPases and contain an N-terminal auto-inhibitory domain upstream of a formin-homology 2 (FH2) domain and an activatory domain (DAD; Figure 3): the FH2/DAD domains interact with actin. The auto-inhibitory domain of mammalian Diaphonous (mDia) is composed of five ARM-repeats (Figure 3) identified by solution of their crystal structure [6,7]. The ARM-repeat domain binds to both RhoA and to the activatory domain of mDia, suggesting a mechanism whereby RhoA binding to the ARM-repeat domain exposes the FH2/DAD, enabling it to interact with actin [6,7]. mDia isoform-specific residues within the ARM-repeats determine the choice of small GTPase binding partner [73]. Interestingly, mDia binds to APC to regulate microtubule stabilisation, but this does not require the ARM-repeats of mDia [74].

Flagellar ARM-repeat proteins

A subset of ARM-repeat proteins is specifically associated with a conserved eukaryotic microtubule-containing structure, the flagellum (or cilium). The well-studied PF16

Box 3. Going green: algal and plant ARM-repeat proteins

Our searches show that even tiny green algae have a healthy complement of ARM-repeat proteins compared with other unicells with similar sized genomes. Higher plants have undergone a larger expansion in ARM-repeat proteins than any other lineage [94]. The relatively large number of green algal ARM-proteins partly explains the large plant ARM-repeat family, because many green algal proteins have higher plant homologues, indicating direct ancestral relations (Figure 3; Figure II). The majority of ARM proteins in green algae have land plant relatives; however, we have identified several groups of ARM-repeat proteins that are green algal-specific. Conversely, land plants possess unique domain combinations acquired as the plant lineage evolved [96]. Importantly, some algal ARM proteins have animal and/or Chromoalveolate homologues but no land plant homologues (Figure 3), emphasising that some ancient cellular functions are conserved across kingdoms but have been lost in the land plant lineage. *Chlamydomonas* is a well-established model organism for genetic and cell biological studies: it is likely that molecular genetic methods will rapidly develop for other algae with sequenced genomes, enabling further study of 'green' ARM-protein evolution.

RING-ARM proteins: plant ubiquitin ligase precursors?

The RING (Really Interesting New Gene) domain is structurally related to, and might have given rise to, the U-box. Both domains are a hallmark of E3 ubiquitin ligases [137] and are found widely across eukaryotes. U-box-ARM proteins are unique to land plants [94]. RING-ARM proteins are found in some green algae [62] and a diatom [138]. It is currently unclear whether: (i) U-box-ARM proteins are derived from an algal ancestor and have subsequently undergone massive expansion during land plant evolution, or (ii) whether this is an example of convergent evolution of proteins to achieve the same function, highlighting the importance of the versatility of ARM-repeat domains.

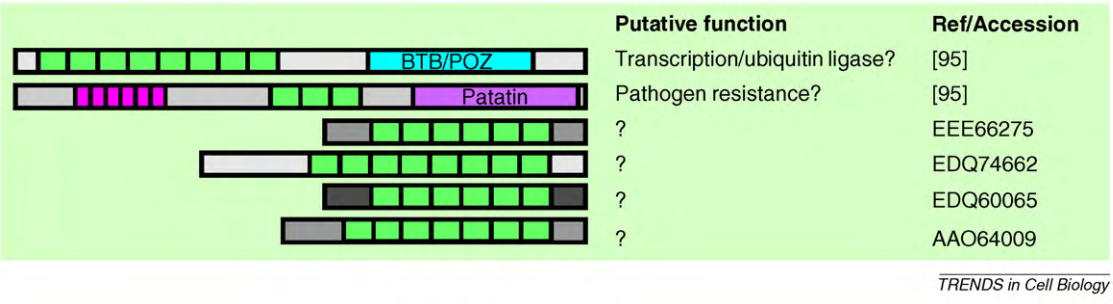


Figure II. Proteins shared between green algae and plants. We have identified six protein families with homologues solely in algae and higher plants. Protein structures are illustrated on the left; green boxes represent ARM repeats; pink boxes represent leucine-rich repeats, other coloured boxes represent domains annotated with a name. Where protein function has been ascertained in at least one species, a review reference is given; where proteins are of unknown function, a representative higher plant GenBank accession number is given.

contains nine ARM-repeats and is evolutionarily ancient, regulating the motility of flagella and cilia in *Chlamydomonas*, mammals and Trypanosomes by stabilizing the central pair of microtubules in the 9+2 arrangement [13,75,76]; Figure 4).

Detection of PF16-related proteins using similarity searches of sequenced genomes correlates well with the presence of a flagellum. PF16 homologues are present in other flagellated green algae, e.g. *Micromonas* sp. ([63]; 75% amino acid identity to *Chlamydomonas* PF16 (GenBank accession number CRU40057)). A predicted protein in the early-branching land plant *Physcomitrella*, which has flagellated male gametes, is 40% identical to *Chlamydomonas* PF16 [77]. We do not detect putative PF16 homologues in non-flagellated red/green algae or higher plants, but proteins related to PF16 exist in various flagellated unicellular chromoalveolate and excavate parasites (Figure 3, Box 2). For example a *Toxoplasma gondii* protein is 64% identical to *Chlamydomonas* PF16. *Plasmodium* sp. possesses a more divergent protein, 37% identical to *Chlamydomonas* PF16. PF16 appears to be absent from diatoms, which although flagellated lack the central pair of microtubules [78].

Analysis of the *Chlamydomonas* flagellar proteome has identified several additional ARM-repeat-containing flagellar proteins [79,80]. We can detect putative homologues of several of these additional proteins (FAP194, RSP8, RSP14, FAP69) in other green algae, chromoalveolates and some animals. Interestingly, the ARM-repeat protein, ARM94, is found in activated sperm tails of the ascidian *Ciona* [81]. ARM94 relatives have not been detected in *Chlamydomonas* [79,80], implicating ARM-repeat proteins in organism-specific modes of flagellar regulation as well as core functions.

Human ARM-repeat-containing protein 6 (ARMC6) interacts with FUSED kinase (see below) and thus might function in flagella ([82]; Figures 3 and 4). ARMC6 probably has multiple roles, being part of the clathrin coat assembly complex, and binding to a WD40 repeat-containing protein (WDR8) involved in bone development [82,83]. We can detect putative ARMC6 homologues in green algae, plants, animals, some chromoalveolates and some excavates (Figure 3): determining ARMC6 function in diverse systems should reveal its conserved and species-specific functions.

ARM-repeat-containing kinases with possible cytoskeletal functions

Several subfamilies of ARM-repeat proteins also contain a kinase domain; their functions will be discussed in this section. Adding the ARM protein–protein interaction domain to a kinase could have strong evolutionary benefit. Since some kinases function within a protein complex to stabilize interactions with target proteins (e.g. [16]), adding a protein–protein interaction domain such as ARM-repeats to a kinase could potentially enable stable interactions with target proteins without the need for additional protein partners.

The Fused serine/threonine kinase was originally discovered in *Drosophila*, as part of the Hedgehog (Hh) developmental signalling pathway. Most Fused kinases consist of an N-terminal kinase domain with C-terminal

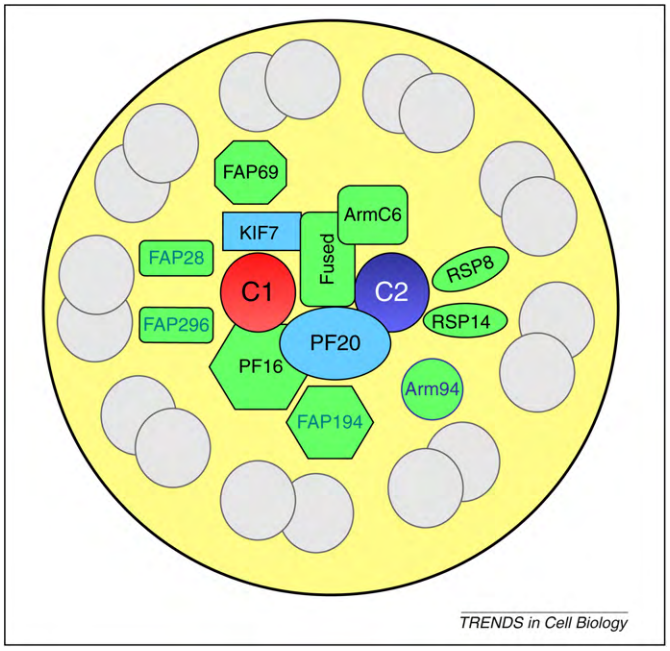


Figure 4. Flagellar ARM-repeat proteins. A cross-section through a eukaryotic flagellum is shown, with the outer microtubule doublets in grey and the inner central pair microtubules labelled C1 (red) and C2 (dark blue). ARM-repeat proteins are shown in green; non-ARM proteins in light blue. PF16 is required for the stability of the C1 microtubule, whereas PF20 is required for the formation of both central pair microtubules. PF20 interacts with the Fused kinase, which also interacts with ARMC6 and the kinesin motor protein KIF7. RSP8 and RSP14 were identified as radial spoke proteins, presumably functioning between the central pair and the outer microtubules. FAP194 is a protein of unknown function that is closely related to PF16. FAP69, FAP28 and FAP296 are ARM-repeat flagellar-associated proteins of unknown function. Proteins with black outlines and labels are widely conserved across kingdoms; those with green outlines and labels are plant/algal (FAP194) or *Chlamydomonas*-specific (FAP296, FAP28); that with a blue outline and label (FAP94) has only been identified in *Ciona*.

ARM-repeats: the ARM-repeats are either lost or highly divergent in *Drosophila* [84,85] but are conserved in plants, some green algae and protists ([78,84,85]; Figure 3). Although vertebrate Fused homologues can regulate Hh signalling in certain cellular contexts [78,86], mouse *fused* knock-outs have normal Hh signalling, but defective ciliated tissues [87]. Mammalian Fused is required to build motile (9+2) cilia and interacts with KIF7, a kinesin that associates with the central pair of microtubules [78]. Interestingly, Fused also interacts with the WD40 protein PF20, which in turn interacts with the ARM-repeat protein PF16, suggesting a very direct regulatory role for multiple ARM-repeat proteins in flagellar central pair construction (Figure 4).

Fused kinases are present in non-flagellated organisms, suggesting they have additional cytoskeletal functions. The *Arabidopsis* Fused homologue, TIO, controls the final stages of cell division [84]. TIO protein localizes to the phragmoplast, a plant-specific structure containing both microtubules and actin, which controls cell wall deposition and hence creation of two daughter cells. *Drosophila fused* mutants have defective female germline mitosis and develop ovarian tumours [88], suggesting that a general role for Fused in eukaryotic cell division might exist. Targeted knockouts of Fused homologues in a range of unicells would be highly informative in this respect. In *Dictyostelium*, no role for a Fused homologue in cell division is documented [85]. However, *Dictyostelium* Fused is required for chemotaxis (directed cell movement) and correct cell polarity, and localises to the microtubule cytoskeleton via its ARM-repeat domain [85].

The kinase-ARM protein domain combination has probably arisen more than once during eukaryotic evolution. ARM repeat-containing kinases related to Mitogen Activated Protein-kinase-kinase-kinase-7 (MAPKKK7) control cell division in *Arabidopsis* pollen, *Dictyostelium* amoebae and *Schizosaccharomyces pombe* [89–91]. Curiously, MAPKKK7-ARM proteins affect only one specific type of cell division in plants, in contrast to their more general role in fission yeast and *Dictyostelium*. Perhaps MAPKKK-ARM proteins only control haploid cell divisions. Testing this hypothesis requires determining the functions of MAPKKK7-ARM proteins in other unicellular systems: we can detect putative homologues in green algae and chromoalveolates (Figure 3).

Never In Mitosis A (NIMA)-kinases regulate cell division in yeasts and mammals [92]. An *Arabidopsis* NIMA-kinase, NEK6, binds to an ARM-repeat-containing protein, ARK1, to regulate cell morphogenesis via the microtubule cytoskeleton [93]. Our searches reveal proteins containing both a NIMA kinase domain and ARM-repeats in some green algae, some chromoalveolates, *Monosiga*, and animals (Figure 3). Generating NIMA-ARM loss-of-function mutants in both unicells and multicellular species will determine whether these proteins have conserved cytoskeletal and cell division functions.

ARM-repeat proteins in protein degradation

A cellular function well-served by ARM-proteins is that of targeted ubiquitination and subsequent degradation of proteins. Many ARM-repeat proteins act as E3 ubiquitin ligases, which interact with, and transfer ubiquitin directly to, a target protein [54]. ARM-repeat-containing ubiquitin ligases are most prevalent in land plants and have been reviewed elsewhere [94,95]; (Box 3). As mentioned previously, Aardvark and Arabidillo proteins both contain F-boxes (Figure 3), although whether they are true ubiquitin ligases is unclear.

ARM-HECT proteins

ARM-HECT proteins are the most widespread ARM-repeat-containing ubiquitin ligases in eukaryotes. Their functions have been characterized in *Arabidopsis* and *S. cerevisiae* but we can detect putative homologues in green and red algae, animals, and *Dictyostelium* (Figure 3). *Arabidopsis* KAKTUS/UPL3 controls DNA replication and regulates trichome (leaf hair) branching in concert with the plant hormone gibberellin [96]. The yeast ARM-HECT protein Ubiquitin-Fusion-Degradation 4 (UFD4) is a ubiquitin ligase during DNA repair [97]. Establishing the function of ARM-HECT proteins in other species will determine whether these proteins have any conserved substrates.

Other UFD ARM-repeat proteins

S. cerevisiae UFD3 has 6 ARM-repeats C-terminal to a WD40 domain. The UFD3 ARM-repeats would not have been identified without structural analysis [11]. As with UFD4, UFD3 is an integral part of the yeast protein degradation process and our searches indicate that

putative ARM-repeat-containing UFD3 homologues exist in animals, higher plants, some chromoalveolates and some excavates (Figure 3). Curiously, another ubiquitin ligase in the yeast ubiquitin-degradation pathway, the U-box containing protein UFD2, contains an 'uneven' structure related to ARM-repeats, indicating the prevalence of this structure in protein degradation processes [10].

ARMC8: a multifunctional degradation and cytoskeletal regulator?

ARM-repeat-containing protein 8 (ARMC8) is part of the conserved C-terminal to LisH motif (CTLH) complex. The CLTH complex is involved in both proteasome/polyubiquitin-dependent protein degradation and vacuole/lysosome-mediated protein degradation (via monoubiquitination and endocytosis of target proteins) in both yeast and human cells [98–100]. Substrates include yeast fructose-1,6-bisphosphatase (in response to glucose availability) and human α-catenin [98,99]. In human cells, the ARM-repeats of ARMC8 potentiate the interaction of membrane/receptor proteins with a signalling adapter, allowing receptor mono-ubiquitination and lysosomal degradation [100], perhaps by providing a scaffold to assemble stable protein–protein interactions.

Proteins in the CLTH complex (apart from ARMC8) possess a LisH/CLTH domain, which is found in proteins that regulate microtubule dynamics and cell division, suggesting that ARMC8 might also have a cytoskeletal function. Accordingly, the CLTH complex associates with microtubules and CLTH domains bind to ARMC8 [98].

We can detect possible ARMC8 homologues in *Dictyostelium*, green algae and land plants (Figure 3), which are 20–24% identical to human ARMC8 (GenBank accession BC032661). Given the key importance of protein degradation in plant development and environmental responses, analysis of plant and algal ARMC8-like proteins could yield particularly interesting data.

New research avenues?

Our searches identify some ARM-repeat proteins about which relatively little is known, which show cross-kingdom conservation of domain combinations, suggesting that these protein families have important functions that should be the target of further analysis in a variety of species.

DUF1716-ARM proteins

The human protein CTNNBL1 (Catenin Beta-Like 1) is an evolutionarily conserved and multifunctional nuclear protein that contains a DUF1716 domain and ARM-repeats [101]. CTNNBL1 interacts with the splicing machinery and also controls class-switching (hypermutation) to generate immunoglobulin diversity [102]. CTNNBL1 also interacts with Osteopontin (OPN), which has been implicated in bone formation and integrin-linked cell adhesion during embryo development and tumour formation [103]. We can detect CTNNBL-like proteins in *Arabidopsis* (38% identical to human CTNNBL1; GenBank accession AL023804), the green alga *Ostreococcus tauri* (33% identity) and *Plasmodium* (24% identity) (Figure 3). Clearly CTNNBL1 has evolved animal-specific functions in immunity and bone formation; ascertaining its functions in unicells will uncover possible ancestral cellular functions.

RCD1-related proteins

Regulator of Cell Differentiation 1 (RCD1) proteins have 6 ARM-repeats and form dimers [104]. They are transcriptional cofactors that might act as cellular stress sensors in yeast and as components of a steroid hormone-induced cell differentiation pathway in mammals; they also degrade RNA [105–107]. We detect putative RCD1 homologues in most eukaryotic organisms tested (Figure 3), suggesting that they could regulate cell differentiation in response to environmental stimuli across kingdoms: this is an interesting area for future investigation.

ARM-TIR-SAM proteins: conserved adapter proteins in defence signalling?

The Toll-Interleukin-Receptor (TIR) domain is found in proteins that mediate signalling in the immune system [108]. The Sterile Alpha Motif (SAM) domain was originally characterized in *S. cerevisiae* pheromone signalling [109]. Animal ARM-TIR-SAM proteins are defence-signalling adaptors. One such protein, mammalian SARM, functions in stress-induced neuronal cell death and could modulate the inflammatory response in a localized area to prevent damage. Interestingly, the ARM-domain associates with mitochondria [110,111]. The *C. elegans* SARM orthologue mediates innate immunity and pathogen resistance in the worm, in part by controlling the production of antimicrobial peptides, and also has a non-immune function localising odourant receptors to the correct olfactory neurons [112].

We can identify ARM-TIR-SAM proteins in unicellular green algae (Figure 3). Because other SAM proteins mediate signals between two cells or between cells and their environment, it will be exciting to ascertain whether ARM-TIR-SAM proteins play a defensive role in algae.

TPR-ARM-repeat proteins: molecular chaperones?

Proteins containing both Tetratricopeptide (TPR)-repeats and ARM-repeats appear to have arisen several times in eukaryotes. TPR repeats form versatile protein–protein interaction domains, and a subset of TPR proteins act as molecular chaperones [113]. The best-characterized TPR-ARM protein is animal UNC45, which regulates the assembly of muscle myosin filaments [114]. Mutations in one human *UNC45* isoform are associated with cardiac myopathies, whereas UNC45-related proteins in unicellular fungi are required for cell division and vesicle transport [114]. A second TPR-ARM protein present in animals is TTC12/TPARM, which is hypermethylated in certain leukaemias and could be associated with alcohol- and drug-dependence in humans [115,116].

We can detect the TPR-ARM or ARM-TPR domain combination in several taxonomic groups (Figure 3; Box 2). We believe that the TPR-ARM structure has arisen several times, because different TPR-ARM proteins have very little amino acid similarity with one another.

ARM-repeat proteins: future directions and open questions

ARM-repeat proteins are found throughout eukaryotes and are evolutionarily ancient (Box 1). The versatile ARM-repeat structure enables diverse essential cellular

functions. The recent and ongoing sequencing of genomes from throughout the eukaryotic tree of life means that it is timely for us to extend our functional studies of these exciting proteins to new systems in addition to the animals, fungi and plants where the majority of ARM-protein characterisation has taken place so far. This will enhance our understanding of how this multifunctional protein family evolved its key cell biological roles.

Acknowledgements

We thank Tony Holder for useful discussions and Laura Moody and Younousse Saidi for critical reading of the manuscript. We acknowledge the Leverhulme Trust (Research Project Grant F/00094/BA - JC) and the Medical Research Council (MRC Investigator grant G0900109/90687 - RT) for support. We thank anonymous reviewers for helpful comments. We apologise to those whose work is only cited in review form owing to lack of space.

References

1 Peifer, M. *et al.* (1994) A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell* 76, 789–791
2 Choi, H.J. and Weis, W.I. (2005) Structure of the armadillo repeat domain of plakophilin 1. *J. Mol. Biol.* 346, 367–376
3 Conti, E. *et al.* (1998) Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell* 94, 193–204
4 Huber, A.H. *et al.* (1997) Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* 90, 871–882
5 Liu, J. *et al.* (2008) The C. elegans SYS-1 protein is a bona fide beta-catenin. *Dev. Cell* 14, 751–761
6 Otomo, T. *et al.* (2005) Structural basis of Rho GTPase-mediated activation of the formin mDia1. *Mol. Cell* 18, 273–281
7 Rose, R. *et al.* (2005) Structural and mechanistic insights into the interaction between Rho and mammalian Dia. *Nature* 435, 513–518
8 Shomura, Y. *et al.* (2005) Regulation of Hsp70 function by HspBP1: structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange. *Mol. Cell* 17, 367–379
9 Striegl, H. *et al.* (2010) Armadillo motifs involved in vesicular transport. *PLoS One* 5, e8991
10 Tu, D. *et al.* (2007) Inaugural article: structure and function of the yeast U-box-containing ubiquitin ligase Ufd2p. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15599–15606
11 Zhao, G. *et al.* (2009) An Armadillo motif in Ufd3 interacts with Cdc48 and is involved in ubiquitin homeostasis and protein degradation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16197–16202
12 Kidd, A.R. *et al.* (2005) A beta-catenin identified by functional rather than sequence criteria and its role in Wnt/MAPK signaling. *Cell* 121, 761–772
13 Coates, J.C. (2003) Armadillo repeat proteins: beyond the animal kingdom. *Trends Cell Biol.* 13, 463–471
14 Kippert, F. and Gerloff, D.L. (2009) Highly sensitive detection of individual HEAT and ARM repeats with HHpred and COACH. *PLoS One* 4, e7148
15 Cadigan, K.M. and Peifer, M. (2009) Wnt signaling from development to disease: insights from model systems. *Cold Spring Harb. Perspect. Biol.* 1, a002881
16 MacDonald, B.T. *et al.* (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* 17, 9–26
17 Xu, W. and Kimelman, D. (2007) Mechanistic insights from structural studies of beta-catenin and its binding partners. *J. Cell Sci.* 120, 3337–3344
18 Gates, J. and Peifer, M. (2005) Can 1000 reviews be wrong? Actin, alpha-catenin, and adherens junctions. *Cell* 123, 769–772
19 Bahmanyar, S. *et al.* (2008) beta-catenin is a Nek2 substrate involved in centrosome separation. *Genes Dev.* 22, 91–105
20 Hadjihannas, M.V. *et al.* (2010) Conductin/axin2 and Wnt signalling regulates centrosome cohesion. *EMBO Rep.* 11, 317–324
21 Huang, P. *et al.* (2007) A novel role of phospho-beta-catenin in microtubule regrowth at centrosome. *Oncogene* 26, 4357–4371
22 Aoki, K. and Taketo, M.M. (2007) Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. *J Cell Sci* 120, 3327–3335
23 Major, M.B. *et al.* (2007) Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. *Science* 316, 1043–1046
24 Chitalia, V.C. *et al.* (2008) Jade-1 inhibits Wnt signalling by ubiquitylating beta-catenin and mediates Wnt pathway inhibition by pVHL. *Nat. Cell Biol.* 10, 1208–1216
25 Dimitrova, Y.N. *et al.* (2010) Direct ubiquitination of β-catenin by siah-1 and regulation by the exchange factor TBL1. *J. Biol. Chem.* 285, 13507–13516
26 Broun, M. *et al.* (2005) Formation of the head organizer in hydra involves the canonical Wnt pathway. *Development* 132, 2907–2916
27 Lapebie, P. *et al.* (2009) WNT/beta-catenin signalling and epithelial patterning in the homoscleromorph sponge *Oscarella*. *PLoS One* 4, e5823
28 Nichols, S.A. *et al.* (2006) Early evolution of animal cell signaling and adhesion genes. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12451–12456
29 Schneider, S.Q. *et al.* (2003) Protein evolution: structure-function relationships of the oncogene beta-catenin in the evolution of multicellular animals. *J. Exp. Zool. B Mol. Dev. Evol.* 295, 25–44
30 Bellipanni, G. *et al.* (2006) Essential and opposing roles of zebrafish beta-catenins in the formation of dorsal axial structures and neurectoderm. *Development* 133, 1299–1309
31 Phillips, B.T. and Kimble, J. (2009) A new look at TCF and beta-catenin through the lens of a divergent C. elegans Wnt pathway. *Dev. Cell* 17, 27–34
32 Chidgey, M. and Dawson, C. (2007) Desmosomes: a role in cancer? *Br. J. Cancer* 96, 1783–1787
33 Martin, E.D. *et al.* (2009) Plakoglobin has both structural and signalling roles in zebrafish development. *Dev. Biol.* 327, 83–96
34 Schmidt, A. and Koch, P.J. (2007) Desmosomes: just cell adhesion or is there more? *Cell Adh. Migr.* 1, 28–32
35 Hatzfeld, M. (2005) The p120 family of cell adhesion molecules. *Eur. J. Cell Biol.* 84, 205–214
36 Franz, C.M. and Ridley, A.J. (2004) p120 catenin associates with microtubules: inverse relationship between microtubule binding and Rho GTPase regulation. *J. Biol. Chem.* 279, 6588–6594
37 Daniel, J.M. (2007) Dancing in and out of the nucleus: p120(ctn) and the transcription factor Kaiso. *Biochim. Biophys. Acta* 1773, 59–68
38 Kim, S.W. *et al.* (2004) Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nat. Cell Biol.* 6, 1212–1220
39 Park, J.I. *et al.* (2005) Kaiso/p120-catenin and TCF/beta-catenin complexes coordinately regulate canonical Wnt gene targets. *Dev. Cell* 8, 843–854
40 Oh, M. *et al.* (2009) GSK-3 phosphorylates delta-catenin and negatively regulates its stability via ubiquitination/proteosome-mediated proteolysis. *J. Biol. Chem.* 284, 28579–28589
41 Iioka, H. *et al.* (2009) Kaiso is a bimodal modulator for Wnt/beta-catenin signaling. *FEBS Lett.* 583, 627–632
42 Park, J.I. *et al.* (2006) Frodo links Dishevelled to the p120-catenin/Kaiso pathway: distinct catenin subfamilies promote Wnt signals. *Dev. Cell* 11, 683–695
43 Ferber, E.C. *et al.* (2008) A role for the cleaved cytoplasmic domain of E-cadherin in the nucleus. *J. Biol. Chem.* 283, 12691–12700
44 Bass-Zubek, A.E. *et al.* (2009) Plakophilins: multifunctional scaffolds for adhesion and signaling. *Curr. Opin. Cell Biol.* 21, 708–716
45 Wolf, A. *et al.* (2010) Plakophilin 1 stimulates translation by promoting eIF4A1 activity. *J. Cell Biol.* 188, 463–471
46 Harwood, A. and Coates, J.C. (2004) A prehistory of cell adhesion. *Curr. Opin. Cell Biol.* 16, 470–476
47 Wang, Y.X. *et al.* (1998) Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole. *J. Cell Biol.* 140, 1063–1074
48 Tang, F. *et al.* (2006) Vac8p, an armadillo repeat protein, coordinates vacuole inheritance with multiple vacuolar processes. *Traffic* 7, 1368–1377
49 Barelle, C.J. *et al.* (2006) Candida albicans VAC8 is required for vacuolar inheritance and normal hyphal branching. *Eukaryot. Cell* 5, 359–367
50 Veses, V. *et al.* (2009) Vacuole inheritance regulates cell size and branching frequency of *Candida albicans* hyphae. *Mol. Microbiol.* 71, 505–519

51 Carr, M. *et al.* (2008) Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16641–16646
52 Abedin, M. and King, N. (2008) The premetazoan ancestry of cadherins. *Science* 319, 946–948
53 Coates, J.C. *et al.* (2006) Armadillo-related proteins promote lateral root development in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1621–1626
54 Petroski, M.D. and Deshaies, R.J. (2005) Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* 6, 9–20
55 Gebert, M. *et al.* (2008) F-actin organization and pollen tube tip growth in *Arabidopsis* are dependent on the gametophyte-specific Armadillo repeat protein ARO1. *Plant Cell* 20, 2798–2814
56 Manning, B.D. and Snyder, M. (2000) Drivers and passengers wanted! The role of kinesin-associated proteins. *Trends Cell Biol.* 10, 281–289
57 Haraguchi, K. *et al.* (2006) Role of the kinesin-2 family protein, KIF3, during mitosis. *J. Biol. Chem.* 281, 4094–4099
58 Choi, J. *et al.* (2008) Kinesin superfamily-associated protein 3 is preferentially expressed in glutamatergic neurons and contributes to the excitatory control of female puberty. *Endocrinology* 149, 6146–6156
59 Tateno, M. *et al.* (2009) Mutant SOD1 impairs axonal transport of choline acetyltransferase and acetylcholine release by sequestering KAP3. *Hum. Mol. Genet.* 18, 942–955
60 Landers, J.E. *et al.* (2009) Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9004–9009
61 King, N. *et al.* (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788
62 Merchant, S.S. *et al.* (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318, 245–250
63 Worden, A.Z. *et al.* (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324, 268–272
64 Armbrust, E.V. *et al.* (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306, 79–86
65 Aurrecoechea, C. *et al.* (2009) GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* 37, D526–D530
66 Aury, J.M. *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444, 171–178
67 Coyne, R.S. *et al.* (2008) Refined annotation and assembly of the *Tetrahymena thermophila* genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. *BMC Genomics* 9, 562
68 Haas, B.J. *et al.* (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461, 393–398
69 Tyler, B.M. *et al.* (2006) Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313, 1261–1266
70 Eichinger, L. *et al.* (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435, 43–57
71 Chesarone, M.A. *et al.* (2010) Unleashing formins to remodel the actin and microtubule cytoskeletons. *Nat. Rev. Mol. Cell Biol.* 11, 62–74
72 Thomas, C. *et al.* (2009) Actin bundling in plants. *Cell Motil. Cytoskeleton* 66, 940–957
73 Lammers, M. *et al.* (2008) Specificity of interactions between mDia isoforms and Rho proteins. *J. Biol. Chem.* 283, 35236–35246
74 Wen, Y. *et al.* (2004) EB1 and APC bind to mDia to stabilize microtubules downstream of Rho and promote cell migration. *Nat. Cell Biol.* 6, 820–830
75 Branche, C. *et al.* (2006) Conserved and specific functions of axoneme components in trypanosome motility. *J. Cell Sci.* 119, 3443–3455
76 Ralston, K.S. *et al.* (2006) Flagellar motility contributes to cytokinesis in *Trypanosoma brucei* and is modulated by an evolutionarily conserved dynein regulatory system. *Eukaryot. Cell* 5, 696–711
77 Rensing, S.A. *et al.* (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69
78 Wilson, C.W. *et al.* (2009) Fused has evolved divergent roles in vertebrate Hedgehog signalling and motile ciliogenesis. *Nature* 459, 98–102
79 Pazour, G.J. *et al.* (2005) Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* 170, 103–113
80 Yang, P. *et al.* (2006) Radial spoke proteins of *Chlamydomonas* flagella. *J. Cell Sci.* 119, 1165–1174
81 Hozumi, A. *et al.* (2008) Molecular characterization of axonemal proteins and signaling molecules responsible for chemoattractant-induced sperm activation in *Ciona intestinalis*. *Cell Motil. Cytoskeleton* 65, 249–267
82 Ewing, R.M. *et al.* (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* 3, 89
83 Koshizuka, Y. *et al.* (2001) Isolation, characterization, and mapping of the mouse and human WDR8 genes, members of a novel WD-repeat gene family. *Genomics* 72, 252–259
84 Oh, S.A. *et al.* (2005) A divergent cellular role for the FUSED kinase family in the plant-specific cytokinetic phragmoplast. *Curr. Biol.* 15, 2107–2111
85 Tang, L. *et al.* (2008) tsunami, the *Dictyostelium* homolog of the Fused kinase, is required for polarization and chemotaxis. *Genes Dev.* 22, 2278–2290
86 Maloveryan, A. *et al.* (2007) A possible role of mouse Fused (STK36) in Hedgehog signaling and Gli transcription factor regulation. *J. Cell Commun. Signal.* 1, 165–173
87 Merchant, M. *et al.* (2005) Loss of the serine/threonine kinase fused results in postnatal growth defects and lethality due to progressive hydrocephalus. *Mol. Cell. Biol.* 25, 7054–7068
88 Narbonne-Reveau, K. *et al.* (2006) fused regulates germline cyst mitosis and differentiation during *Drosophila* oogenesis. *Mech. Dev.* 123, 197–209
89 Chaiwongsar, S. *et al.* (2006) The protein kinase genes MAP3K epsilon 1 and MAP3K epsilon 2 are required for pollen viability in Arabidopsis thaliana. *Plant J.* 48, 193–205
90 Fankhauser, C. and Simanis, V. (1994) The cdc7 protein kinase is a dosage dependent regulator of septum formation in fission yeast. *EMBO J.* 13, 3011–3019
91 Muller-Taubenberger, A. *et al.* (2009) The STE group kinase SepA controls cleavage furrow formation in *Dictyostelium*. *Cell Motil. Cytoskeleton* 66, 929–939
92 O'Connell, M.J. *et al.* (2003) Never say never. The NIMA-related protein kinases in mitotic control. *Trends Cell Biol.* 13, 221–228
93 Sakai, T. *et al.* (2008) Armadillo repeat-containing kinesins and a NIMA-related kinase are required for epidermal-cell morphogenesis in Arabidopsis. *Plant J.* 53, 157–171
94 Samuel, M.A. *et al.* (2006) Multifunctional arm repeat domains in plants. *Int. Rev. Cytol.* 253, 1–26
95 Yee, D. and Goring, D.R. (2009) The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. *J. Exp. Bot.* 60, 1109–1121
96 Coates, J.C. (2008) Armadillo repeat proteins: versatile regulators of plant development and signalling. In *Plant Growth Signalling* (Bogre, L. and Beemster, G., eds), pp. 299–314, Springer
97 Ju, D. and Xie, Y. (2006) A synthetic defect in protein degradation caused by loss of Ufd4 and Rad23. *Biochem. Biophys. Res. Commun.* 341, 648–652
98 Kobayashi, N. *et al.* (2007) RanBPM, Muskelin, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8alpha and ARMC8beta are components of the CTLH complex. *Gene* 396, 236–247
99 Suzuki, T. *et al.* (2008) Proteasome-dependent degradation of alpha-catenin is regulated by interaction with ARMc8alpha. *Biochem. J.* 411, 581–591
100 Tomaru, K. *et al.* (2010) Armadillo repeat containing 8alpha binds to HRS and promotes HRS interaction with ubiquitinated proteins. *Open Biochem. J.* 4, 1–8
101 Jabbour, L. *et al.* (2003) Sequence, gene structure, and expression pattern of CTNNBL1, a minor-class intron-containing gene – evidence for a role in apoptosis. *Genomics* 81, 292–303
102 Conticello, S.G. *et al.* (2008) Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. *Mol. Cell* 31, 474–484
103 El-Tanani, M.K. (2008) Role of osteopontin in cellular signaling and metastatic phenotype. *Front. Biosci.* 13, 4276–4284
104 Garces, R.G. *et al.* (2007) Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties. *Protein Sci.* 16, 176–188

105 Garapaty, S. *et al.* (2008) Components of the CCR4-NOT complex function as nuclear hormone receptor coactivators via association with the NRC-interacting Factor NIF-1. *J. Biol. Chem.* 283, 6806–6816
106 Haas, M. *et al.* (2004) c-Myb protein interacts with Rcd-1, a component of the CCR4 transcription mediator complex. *Biochemistry* 43, 8152–8159
107 Hiroi, N. *et al.* (2002) Mammalian Rcd1 is a novel transcriptional cofactor that mediates retinoic acid-induced cell differentiation. *EMBO J.* 21, 5235–5244
108 McGettrick, A.F. and O'Neill, L.A. (2004) The expanding family of MyD88-like adaptors in Toll-like receptor signal transduction. *Mol. Immunol.* 41, 577–582
109 Ramezani-Rad, M. (2003) The role of adaptor protein Ste50-dependent regulation of the MAPKKK Ste11 in multiple signalling pathways of yeast. *Curr. Genet.* 43, 161–170
110 Kenny, E.F. and O'Neill, L.A. (2008) Signalling adaptors used by Toll-like receptors: an update. *Cytokine* 43, 342–349
111 Szretter, K.J. *et al.* (2009) The immune adaptor molecule SARM modulates tumor necrosis factor alpha production and microglia activation in the brainstem and restricts West Nile Virus pathogenesis. *J. Virol.* 83, 9329–9338
112 O'Neill, L.A. (2006) DisSARMing Toll-like receptor signaling. *Nat. Immunol.* 7, 1023–1025
113 D'Andrea, L.D. and Regan, L. (2003) TPR proteins: the versatile helix. *Trends Biochem. Sci.* 28, 655–662
114 Kim, J. *et al.* (2008) Protein quality control gets muscle into shape. *Trends Cell Biol.* 18, 264–272
115 Wattanawaraporn, R. *et al.* (2007) Hypermethylation of TTC12 gene in acute lymphoblastic leukemia. *Leukemia* 21, 2370–2373
116 Yang, B.Z. *et al.* (2008) Haplotypic variants in DRD2, ANKK1, TTC12, and NCAM1 are associated with comorbid alcohol and drug dependence. *Alcohol Clin. Exp. Res.* 32, 2117–2127
117 Mason, D.A. *et al.* (2009) Evolution of the metazoan-specific importin alpha gene family. *J. Mol. Evol.* 68, 351–365
118 Goldfarb, D.S. *et al.* (2004) Importin alpha: a multipurpose nuclear-transport receptor. *Trends Cell Biol.* 14, 505–514
119 Holt, J.E. *et al.* (2007) Regulation of nuclear import during differentiation; the imp alpha gene family and spermatogenesis. *Curr. Genomics* 8, 323–334
120 Tejomurtula, J. *et al.* (2009) Role of importin alpha8, a new member of the importin alpha family of nuclear transport proteins, in early embryonic development in cattle. *Biol. Reprod.* 81, 333–342
121 Ratan, R. *et al.* (2008) *Drosophila* importin alpha1 performs paralog-specific functions essential for gametogenesis. *Genetics* 178, 839–850
122 Bhattacharjee, S. *et al.* (2008) IMPa-4, an *Arabidopsis* importin alpha isoform, is preferentially involved in agrobacterium-mediated plant transformation. *Plant Cell* 20, 2661–2680
123 Palma, K. *et al.* (2005) An importin alpha homolog, MOS6, plays an important role in plant innate immunity. *Curr. Biol.* 15, 1129–1135
124 Mans, B.J. *et al.* (2004) Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3, 1612–1637
125 Aravind, L. *et al.* (2006) Comparative genomics and structural biology of the molecular innovations of eukaryotes. *Curr. Opin. Struct. Biol.* 16, 409–419
126 Fagotto, F. *et al.* (1998) Nuclear localization signal-independent and importin/karyopherin-independent nuclear import of beta-catenin. *Curr. Biol.* 8, 181–190
127 Schatz, C.A. *et al.* (2003) Importin alpha-regulated nucleation of microtubules by TPX2. *EMBO J.* 22, 2060–2070
128 Cavalier-Smith, T. (1993) Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* 57, 953–994
129 Lau, A.O. (2009) An overview of the *Babesia, Plasmodium* and *Theileria* genomes: a comparative perspective. *Mol. Biochem. Parasitol.* 164, 1–8
130 Templeton, T.J. *et al.* (2004) Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res.* 14, 1686–1695
131 Wasmuth, J. *et al.* (2009) The origins of apicomplexan sequence innovation. *Genome Res.* 19, 1202–1213
132 Aurrecoechea, C. *et al.* (2007) ApiDB: integrated resources for the apicomplexan bioinformatics resource center. *Nucleic Acids Res.* 35, D427–D430
133 de Koning-Ward, T.F. *et al.* (2000) The development of genetic tools for dissecting the biology of malaria parasites. *Annu. Rev. Microbiol.* 54, 157–185
134 Frenal, K. and Soldati-Favre, D. (2009) Role of the parasite and host cytoskeleton in apicomplexa parasitism. *Cell Host Microbe* 5, 602–611
135 Briggs, L.J. *et al.* (2004) More than one way to build a flagellum: comparative genomics of parasitic protozoa. *Curr. Biol.* 14, R611–R612
136 Rudzinska, M.A. *et al.* (1983) Sexuality in piroplasms as revealed by electron microscopy in *Babesia microti*. *Proc. Natl. Acad. Sci. U. S. A.* 80, 2966–2970
137 Aravind, L. and Koonin, E.V. (2000) The U box is a modified RING finger – a common domain in ubiquitination. *Curr. Biol.* 10, R132–R134
138 Bowler, C. *et al.* (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244
139 Baker, N.A. *et al.* (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10037–10041
140 DeLano, W.L. (2008) *PyMol*, DeLano Scientific LLC
141 Baldauf, S.L. (2003) The deep roots of eukaryotes. *Science* 300, 1703–1706